Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Biotechnology

Means to optimize protein expression in transgenic plants

Kristian K Ullrich[1], Manuel Hiss[1] and Stefan A Rensing[1,2]

The biotechnological production of proteins is currently achieved via expression systems derived from different lineages. In the past years transgenic plants have proven to be able to compete with bacteria or mammalian cell systems. Gene engineering approaches exist to raise yields by controlling mandatory processes in the course of biopharmaceutical protein production. Here we review and discuss the current status and recent improvements of parameters influencing recombinant protein production in transgenic plants. In particular, this review focuses on the so-called inside (mRNA sequence and structure) and outside factors (host and production system/conditions), which are adjustable and allow to optimize protein production via gene engineering.

Addresses
[1] Plant Cell Biology, Faculty of Biology, University of Marburg, Karl-von-Frisch-Str. 8, D-35043 Marburg, Germany
[2] BIOSS Centre for Biological Signalling Studies, University of Freiburg, Germany

Corresponding author: Rensing, Stefan A (stefan.rensing@biologie.uni-marburg.de)

Current Opinion in Biotechnology 2015, **32**:61–67

This review comes from a themed issue on **Plant biotechnology**

Edited by **Inge Broer** and **George N Skaracis**

For a complete overview see the Issue and the Editorial

Available online 25th November 2014

http://dx.doi.org/10.1016/j.copbio.2014.11.011

Introduction

The potential of transgenic plants to act as alternatives to classical/standard expression systems for protein production has led to an increasing interest in this field of biotechnology. Several studies (reviewed in [1•]) have already highlighted that transgenic plants can produce low-cost biomass and yield high amounts of recombinant proteins. Moreover, they can modify pharmaceutically important proteins in a post-transcriptional manner and some of them can be grown like crops [2]. In 2012 the U.S. Food and Drug Administration (FDA) approved the first plant cell-expressed therapeutic drug in a transgenic carrot cell culture system. The yield of the biotechnologically relevant protein is dependent on the used host system, the transformation method, the targeted cellular compartment, the realized expression level in the host system and by the accumulation status of the protein [1•].

Adjustable parameters for gene engineering can be grouped into inside features (IF; mRNA and protein primary structure and features) and outside features (OF; host system, vector system, culture conditions). IFs and OFs are closely connected and impact each other. For most research projects individual parameter sets are evaluated and improved on a per case basis, therefore a bullet-proof correlation of gene features and their predicted impact on the amount of end-product is still lacking. In the same context it is still under debate exactly which impact the changes in codon usage can have on the final protein abundance. In some cases it apparently does not increase gene expression at all and in other cases up to 1000-fold [1•,3•,4,5]. It is also debated how codon usage effects other features of the underlying transcript like for example RNA folding, that might result in changes of protein expression level [6].

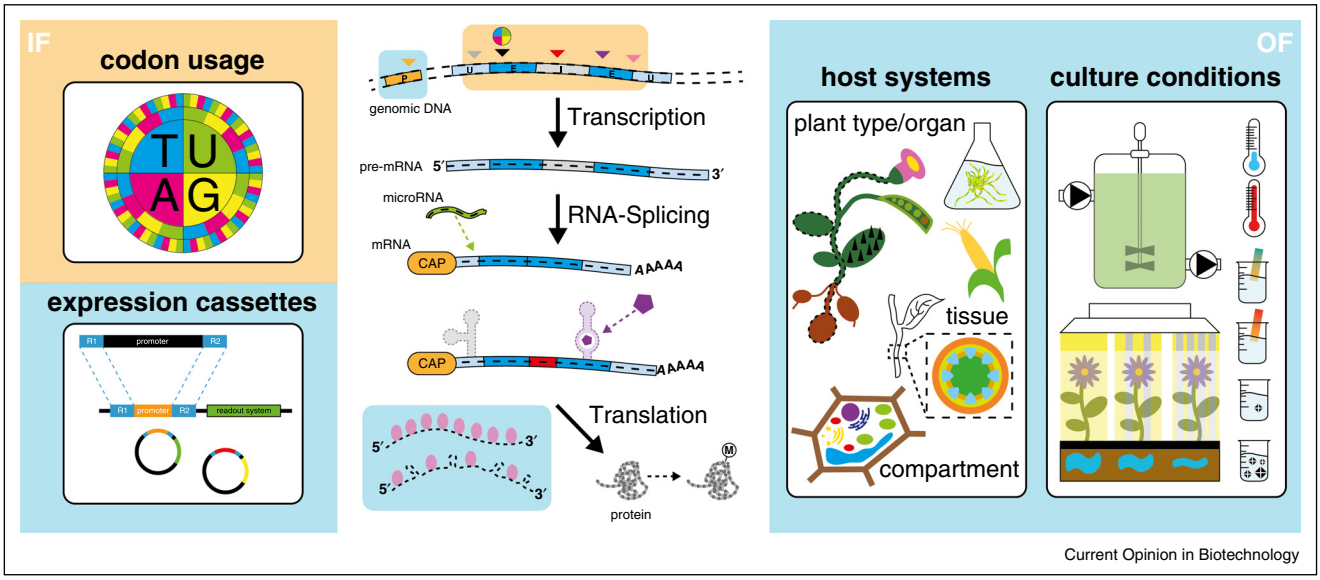
As reviewed in [3•], maximizing heterologous protein expression is a multidimensional optimization problem. To benefit from plant-specific features, one needs to decode the complex nature of multiple factors influencing and controlling the outcome of eukaryotic expression systems. By using new technologies, knowledge gaps are being closed by monitoring the necessary parameters [5]. Recently a study [7] based on a hypothesis about selection constraints and codon bias [8] could link the IF codon bias and OFs like ribosomal dependent translation, tRNA abundance and host evolution in prokaryotes. The question remains whether these findings also apply for eukaryotic expression systems [9].

Next to the already mentioned IF codon bias a diverse set of IF and OF parameters can influence protein expression and should be considered for optimization as highlighted in Figure 1. In this review we give an overview about recent improvements which have been made in optimization, correlation and prediction of these IFs and OFs which altogether influence protein expression in transgenic plants. We will further clarify the importance of experimental design and highlight how new resources and design-rules can be adapted for genetic engineering.

Improving inside features (IFs)

In the process of developing gene engineering, codon bias is an intensively studied parameter influencing protein expression and hence impacts industrial production systems

Figure 1



Schematic overview of inside (beige) and outside features (blue) affecting protein expression. Inside features (IFs) include codon bias, intron sequences and UTRs, while outside features (OFs) include host systems, culture conditions, expression cassettes and promoters. Marked with colored arrows are promoter (orange), UTR (gray), coding sequence (black, violet and salmon) and intron (red).

[10]. Next to a dependence on the OF tRNA abundance and the IF codon bias even altering synonymous sites on the mRNA level (maintaining the original protein sequence) can lead to tremendous effects on transcription level. This can be due to changed folding capacities of the mRNA or by introducing or removing miRNA target sequences or forming new splice acceptor or donor sites [11].

Nucleotide frequencies of mRNA (codon usage and GC content)

One important step in introducing a transgene for protein production comprises altering the codon usage to match the host genome (either be the nuclear genome or the genome of the targeted semi-autonomous organelle). Codon usage data for some plants can be found in the 'codon usage database' (http://www.kazusa.or.jp/codon/ [12]). There are several metrics that can be applied to measure codon usage. In [13] a whole chapter deals with different classification categories, evaluates selected metrics (see Box 1) and concludes that a combination of different metrics may provide the best means for successful alteration of codon composition.

Global views on codon usage can give insights into evolutionary mechanisms that shape the gene sequence and by that influence expression. The nucleotide distributions in seed plants have been reviewed [14] and it was found that the GC-biased gene conversion (gBGC), a recombination-associated process, shapes the nucleotide patterns in more than 200 seed plants. To select between a position-dependent codon bias model and the gBGC model it would be necessary to measure ribosomal speed in selected transcripts. Such technology is already available for bacteria [15].

Reasons for the observed codon bias might be selection for either translational efficiency [16] or translational accuracy, highlighting the interplay between IFs and OFs. Apparently, a mixed model explains best the balanced codon usage found in eukaryotic genomes [17]. However, there are several other factors that influence the choice of codons at certain positions like maintaining mRNA structure [18] or tissue specific expression [19,20]. Akin to [6], previous work could link mRNA stability with the first nucleotides of the coding sequence and further with tRNA abundance and ribosome density during translation [21]. Altering the structure of mRNA can also be used to introduce synthetic riboswitches to induce or repress gene expression via external triggers as reviewed in [22].

These complex interactions are not yet fully understood since only a few studies focus on a multifactorial design to elucidate the effects of changes to the gene sequence on gene expression. One exception has been described [23••] where the effects of different factors in two genes in sugarcane were analyzed. Removal of RNA instability sequences showed the greatest improvement of transgene expression strength. Yet, the authors suggest a change of all features and propose a set of design rules for transgenes.